

A Hybrid Deep Learning and Nature-Inspired Optimization Framework for Accurate and Interpretable Lung Cancer Detection from CT Scans

S Saravana Kumar¹, Shashi Kant Gupta²

Lincoln University College, Malaysia¹

Lincoln University College, Malaysia²

Dept of CSE, PSCMR College of Engineering Technology, Vijayawada¹

saravanakumarmithun@gmail.com, pdf.saravana@lincoln.edu.my, shashigupta@lincoln.edu.my

Abstract

Lung cancer remains one of the leading causes of cancer-related mortality worldwide, primarily due to late diagnosis and complex tumor characteristics. Early detection through medical imaging can significantly improve survival rates. Computed Tomography (CT) scans are widely used for identifying pulmonary nodules; however, manual interpretation by radiologists is time-consuming and susceptible to variability. This study proposes a **hybrid deep learning and nature-inspired optimization framework** for accurate lung cancer detection from CT images. The proposed approach integrates **deep learning-based feature extraction with a nature-inspired optimization algorithm for feature selection**, improving classification performance while maintaining model interpretability. CT images undergo preprocessing to remove noise and enhance relevant structures. A convolutional neural network (CNN) extracts deep features representing nodular characteristics. Subsequently, a nature-inspired optimization algorithm selects the most informative features to reduce dimensionality and improve classification accuracy. The optimized features are fed into a classification model to distinguish between malignant and benign lung nodules. Experimental evaluation demonstrates that the proposed hybrid model improves detection accuracy, sensitivity, and computational efficiency compared with traditional deep learning models. The framework also incorporates interpretability mechanisms that highlight significant image regions influencing the classification outcome, assisting clinicians in understanding model predictions. The results indicate that combining artificial intelligence techniques with nature-inspired optimization can significantly enhance automated lung cancer diagnosis systems.

Keywords: Lung Cancer Detection, Deep Learning, Nature Inspired Optimization, CT Imaging, Medical Image Analysis, Explainable AI.

1. Introduction

Lung cancer is one of the most prevalent and deadly forms of cancer globally. According to recent medical statistics, millions of new lung cancer cases are diagnosed each year, and early detection plays a critical role in improving patient survival rates. Computed Tomography (CT) imaging has become a primary diagnostic tool for detecting pulmonary nodules and identifying possible malignancies. However, the manual analysis of CT scans requires significant expertise and time, and diagnostic outcomes may vary among radiologists.

Artificial intelligence, particularly deep learning, has shown remarkable success in medical image analysis. Convolutional Neural Networks (CNNs) have demonstrated strong performance in detecting patterns and features in medical images, enabling automated detection of abnormalities such as tumors and nodules. Despite their success, deep learning models often generate large numbers of features, some of which may be redundant or irrelevant, potentially affecting classification accuracy and computational efficiency.

To address these limitations, optimization techniques inspired by natural phenomena have been increasingly applied to feature selection problems. Nature-inspired algorithms mimic biological or physical processes such as evolution, swarm intelligence, and natural selection to find optimal solutions in complex search spaces. When integrated with deep learning models, these algorithms can identify the most informative features and eliminate unnecessary ones.

This research proposes a **hybrid framework combining deep learning-based feature extraction with nature-inspired optimization for feature selection**. The objective is to enhance the performance of lung cancer detection systems while reducing computational complexity and improving interpretability. By selecting the most relevant features and integrating them with an effective classification model, the proposed framework aims to deliver accurate and clinically interpretable diagnostic predictions.

2. Methodology

System Architecture

The proposed framework consists of five major stages:

1. Data acquisition and preprocessing
2. Deep learning-based feature extraction
3. Nature-inspired feature selection and optimization
4. Classification of lung nodules
5. Interpretability and clinical integration

This hybrid architecture leverages the strengths of both deep learning and optimization techniques to achieve improved diagnostic performance.

2.1 Data Acquisition and Preprocessing

The dataset used in this study consists of CT scan images containing lung nodules. These images may contain noise, variations in brightness, and irrelevant background information. Therefore, preprocessing plays an essential role in improving the quality of input data.

The preprocessing stage includes:

- **Image normalization** to standardize intensity values.
- **Noise reduction techniques** such as filtering to remove unwanted artifacts.
- **Image resizing and scaling** to ensure consistent input dimensions for the deep learning model.
- **Segmentation techniques** to isolate lung regions from surrounding tissues.

These steps ensure that the input images provide clear and relevant information for feature extraction.

2.2 Feature Extraction Using Deep Learning

Feature extraction is performed using a **Convolutional Neural Network (CNN)** architecture. CNN models are well suited for medical image analysis because they automatically learn hierarchical representations of image features.

The CNN extracts several levels of features:

- **Low-level features** such as edges and textures
- **Mid-level features** representing patterns and shapes
- **High-level features** representing nodular structures and abnormalities

These deep features capture critical information about lung nodules, including their shape, size, and density patterns. The extracted features are then passed to the optimization stage for refinement.

2.3 Feature Selection Using Nature-Inspired Optimization

Although deep learning models generate powerful representations, they often produce a large number of features. Some of these features may not contribute significantly to classification performance. To address this issue, a **nature-inspired optimization algorithm** is applied for feature selection.

Nature-inspired algorithms simulate biological or natural processes to search for optimal solutions. The optimization algorithm evaluates different feature subsets and selects those that maximize classification performance while minimizing redundancy.

The feature selection process involves:

1. Initialization of candidate feature subsets
2. Evaluation using a fitness function based on classification performance
3. Iterative optimization to refine feature combinations
4. Selection of the best-performing feature subset

By reducing dimensionality, the optimization process improves classification accuracy and reduces computational cost.

2.4 Classification Layer

The optimized feature subset is used to train a classification model that distinguishes between **benign and malignant lung nodules**. Various machine learning classifiers can be employed, including:

- Support Vector Machine (SVM)
- Random Forest
- Fully connected neural networks

In this study, the classifier is trained using labeled CT images and optimized features. The model learns decision boundaries that differentiate between cancerous and non-cancerous nodules.

Performance evaluation is conducted using standard classification metrics such as accuracy, precision, recall, and F1-score.

2.5 Interpretability and Clinical Integration

One major challenge in AI-based medical diagnosis is the lack of interpretability. Clinicians require transparent models that explain how predictions are generated. Therefore, the proposed framework includes interpretability mechanisms to highlight the image regions responsible for classification decisions.

Techniques such as **visual attention maps and feature importance analysis** are used to identify critical areas within CT images. These visual explanations help clinicians understand the reasoning behind the model's predictions and increase trust in AI-assisted diagnosis systems.

3. Results and Analysis

3.1 Classification Metrics

The performance of the proposed hybrid framework is evaluated using several classification metrics:

- **Accuracy:** proportion of correctly classified samples
- **Precision:** proportion of correctly predicted positive samples
- **Recall (Sensitivity):** ability to detect actual cancer cases
- **F1-score:** harmonic mean of precision and recall

Experimental results demonstrate that the hybrid approach achieves higher accuracy and improved sensitivity compared with baseline models. The optimization-based feature selection significantly enhances the classifier's ability to distinguish between malignant and benign nodules.

3.2 Computational Metrics

In addition to classification performance, computational efficiency is also evaluated. Feature selection reduces the dimensionality of the feature space, resulting in:

- Reduced training time
- Lower memory consumption
- Improved model stability

The optimized model requires fewer computational resources while maintaining high detection accuracy, making it suitable for real-world clinical applications.

3.3 Comparative Results

The proposed framework is compared with conventional deep learning models that do not use feature optimization. Experimental findings indicate that the hybrid model outperforms existing approaches in terms of both accuracy and efficiency.

Key observations include:

- Improved classification accuracy for lung cancer detection
- Reduced feature redundancy through optimization
- Enhanced interpretability through visualization techniques

These results demonstrate that integrating nature-inspired optimization with deep learning provides a robust solution for automated lung cancer diagnosis.

4. Conclusion

This study presented a **hybrid deep learning and nature-inspired optimization framework** for lung cancer detection using CT scan images. The proposed system combines CNN-based feature extraction with optimization-based feature selection to enhance classification accuracy and computational efficiency.

The experimental results demonstrate that the hybrid model achieves improved performance compared with conventional deep learning approaches. The optimization process effectively selects the most informative features, reducing redundancy and improving diagnostic accuracy. Furthermore, interpretability techniques allow clinicians to visualize the important image regions influencing the model's predictions.

The proposed framework highlights the potential of integrating artificial intelligence and optimization techniques in medical image analysis. Future research may focus on incorporating larger datasets, advanced deep learning architectures, and real-time clinical deployment to further enhance the system's effectiveness in early lung cancer detection.

References

1. Y. Tian, L. Feng, and H. Chen, "Transfer Learning with ResNet for Early Lung Cancer Detection," *IEEE Trans. Med. Imaging*, vol. 42, no. 1, pp. 15–25, 2023.
2. R. Sharma et al., "Feature Selection Using PSO for Lung Cancer Classification," *Comput. Biol. Med.*, vol. 138, p. 104899, 2021.
3. P. Mehra and A. Singh, "Genetic Algorithm-Based Feature Reduction for Lung Cancer," *Expert Syst. Appl.*, vol. 183, p. 115471, 2022.

4. H. Alzubaidi et al., "Ant Colony Optimization in Histopathological Image Classification," *Pattern Recognit.*, vol. 140, p. 109589, 2023.
5. T. Nguyen et al., "Hybrid CNN-GA Model for Lung Cancer Detection," *IEEE Access*, vol. 10, pp. 9845–9855, 2022.
6. D. Patel et al., "LSTM and PSO Hybrid Model for CT Scan Analysis," *Biomed. Signal Process. Control*, vol. 82, p. 104025, 2023.
7. S. Lee and D. Kim, "Dual-Stage ACO-Based Classifier for Lung Nodules," *Artif. Intell. Med.*, vol. 122, p. 102243, 2023.
8. A. Rahman et al., "NSGA-II for Multi-Objective Classification," *Appl. Soft Comput.*, vol. 121, p. 108734, 2022.
9. S. Jha et al., "Firefly Algorithm for Ensemble Classifiers," *Neurocomputing*, vol. 512, pp. 345–354, 2023.
10. B. Zhou et al., "Fuzzy-PSO-Based Medical Image Segmentation," *Signal Process. Image Commun.*, vol. 112, p. 116812, 2024.
11. N. Verma et al., "Comparative Study on SVM, DL, and DL-NIA Models," *J. Med. Syst.*, vol. 47, no. 2, p. 15, 2023.